



(12) **EUROPEAN PATENT APPLICATION**

(21) Application number : **95302109.4**

(51) Int. Cl.⁶ : **H04M 1/27, H04M 3/50**

(22) Date of filing : **29.03.95**

(30) Priority : **06.04.94 US 223810**

(43) Date of publication of application :
11.10.95 Bulletin 95/41

(84) Designated Contracting States :
**AT BE CH DE DK ES FR GB GR IE IT LI LU MC
NL PT SE**

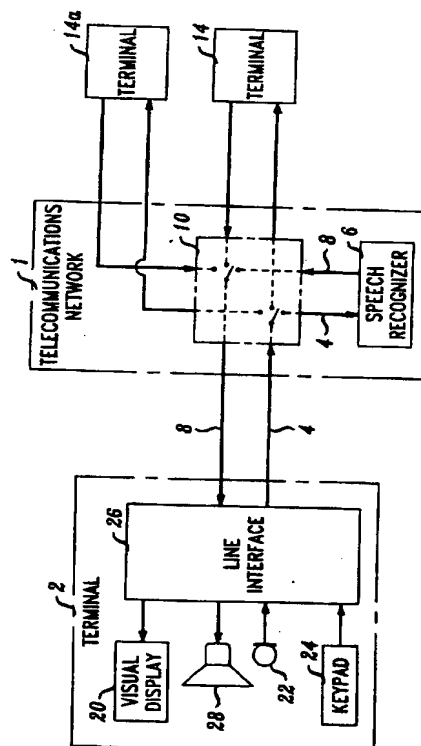
(71) Applicant : **AT & T Corp.**
32 Avenue of the Americas
New York, NY 10013-2412 (US)

(72) Inventor : **Haimi-Cohen, Razel**
30 N. Derby Road,
Springfield
New Jersey 07081 (US)
Inventor : **Reed, Adam Victor**
23 Longfellow Terrace,
Morganville
New Jersey 07751 (US)

(74) Representative : **Watts, Christopher Malcolm**
Kelway, Dr. et al
AT&T (UK) Ltd.
5, Mornington Road
Woodford Green Essex, IG8 0TU (GB)

(54) **Speech recognition system with display for user's confirmation.**

(57) A speech recognizer system for use with a telecommunication network wherein an input signal generated onto the network from a first terminal is directed to a speech recognizer for estimating the verbal content of the input signal. The speech recognizer or associated equipment then directs an estimate of the verbal content as an output signal back to the first terminal, the estimate including one or more approximations of the verbal content of the input signal. At the first terminal the user then confirms a correct estimate, or selects from a plurality of approximations, the verbal content of the input signal.



EP 0 676 882 A2

selection of an approximation.

Preferably, the first terminal 2 has a visual display 20 and the output signal 8 is in a digital format when sent to the first terminal 2 to speed the movement of the output signal 8 on the network to the first terminal 2. Preferably the terminal will have a modem to decode the digital signal for presentation. Alternatively, the output signal 8 will be in DTMF, which can be used with most all current terminal system, that is decoded and presented in visual form at the terminal. The display 20 of the terminal 2 will then decode the signal 8 and visually present it to the user.

At the first terminal 2 the output signal 8 is received and at least a portion thereof, i.e. one approximation, is presented to the user for confirmation. In its preferred embodiment the most probable approximation will be displayed first, followed by the next most probable if the first is not selected or confirmed.

Of course, although a visual display 20 is preferred, the presentation can be audio via a speaker 28, visual or both, with visual alone or in combination with audio being preferred in most instances due to the speed of presentation and reduced need for user alertness where the visual information can stay on a display 20 until the user wishes to remove it, by confirming the correctness or selecting the next approximation.

However, for car phones where the user has his eyes on the road or with telephones that do not have a visual display, audio presentation via speaker 28 is available alone or along with the visual display 20.

In situations where an audio presentation is made, "barge-in" capability is especially important so a user does not need to wait for the end of the audio presentation to confirm a correct approximation or request the next selection. The barge-in feature allows the user to make a confirmation or request another approximation, by depressing a key on the keypad 24 or speaking into the microphone 22 during the presentation, thereby terminating the presentation of the previous approximation without having to listen to the entire presentation.

The preferred visual display 20 can be of any type, including a Caller I.D. where a line or more of alphanumeric text is presented in an LCD display, a P.C. monitor, a CRT display, a vacuum fluorescent display, an LED display, a video telephone, a still image telephone, etc.

In implementing the speech recognition system of the present invention, a communication protocol must be defined for transmitting the output signal 8 from the network 1 to the first terminal 2. Definition of the protocol requires that the variety of possible terminal types and visual displays present in the network be taken into account. Several methods are currently envisioned herein, including a bidirectional protocol, a terminal specific protocol and a unidirectional protocol.

A bidirectional protocol requires that the terminal 2 respond to the network prompt and describe the capabilities of the terminal 2. The network can then direct an output signal 8 to the terminal 2 which matches the capabilities of the terminal 2. For example, if the line interface 26 of the first terminal 2 has a high speed modem, the output 8 will be set faster using the modem protocol. If the terminal 2 is a videophone or still image telephone, the system will generate an output signal 8 comprising a video image for transmission to the visual display 20 terminal 2. If the terminal 2 can display more than one approximation, the estimate may be transmitted by the network for visual display of more than one approximation and prompt the user by synthesized speech, etc., to choose. In the bidirectional protocol a terminal which does not respond to the prompt will be considered to not have any visual display 20 and the output 8 will be in the form of synthesized speech.

With a terminal specific protocol, the network 1 stores a table of the identities of each terminal 2 and utilizes a terminal specific protocol based on the information on the specific terminal. This approach, however, would only be effective in a small network where a network administrator has control over all of the installed terminals.

With a unidirectional protocol the network transmits both a digital feedback for visual display and an audio synthesized speech feedback for audio presentation to the user at the first terminal 2. The format is fixed and the specific terminal can ignore or display the digital feedback for visual display. Of course, this is the most simple protocol, however, it does not allow for customization to specific terminals.

When the presentation is made to the user at the first terminal 2, the user is able to confirm a correct estimate. This includes the ability to indicate that a correct estimate is displayed or request another alternative if additional alternatives are available. If a multi-line display, e.g. a CRT display is used, the confirmation means includes selection means to select from the approximations displayed, to scroll down or bring up a new screen of additional approximations, etc. Such means includes a keypad 24 or a microphone 22 for voice input.

Additionally, the feedback can be augmented with other information resulting from the query of a database with a recognized input or the most closely matching approximations of an input. For example, in a telephone directory application response to an input signal may include an estimate including the most closely matching name or names together with the corresponding telephone numbers. Similarly, in an exchange request the cost of calling each of the approximations can be included. In such applications the confirmation feature can include automatic dialing of the selected approximation or a request for the next screenful of approximations.

most probable alternative presented after said most probable approximation only if the most probable approximation has not been confirmed.

12. A method of error reduction in speech recognition in a telecommunications network comprising the steps of providing an estimate of the content of an input signal placed on the network from a user at first terminal means back to the first terminal means, said estimate comprising more than one approximation of the content of the input signal, and providing confirmation or selection of a correct approximation from the first terminal means onto the network.

5

10

15

20

25

30

35

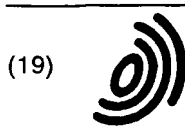
40

45

50

55

5



Europäisches Patentamt
European Patent Office
Office européen des brevets



(11) **EP 0 676 882 A3**

(12) **EUROPEAN PATENT APPLICATION**

(88) Date of publication A3:
06.03.1996 Bulletin 1996/10

(51) Int Cl.⁶: **H04M 1/27, H04M 3/50**

(43) Date of publication A2:
11.10.1995 Bulletin 1995/41

(21) Application number: **95302109.4**

(22) Date of filing: **29.03.1995**

(84) Designated Contracting States:
AT BE CH DE DK ES FR GB GR IE IT LI LU MC NL PT SE

• **Reed, Adam Victor**
New Jersey 07751 (US)

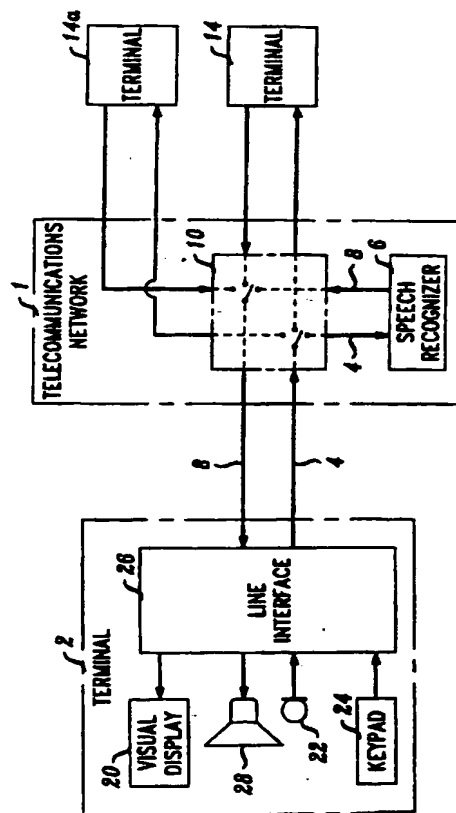
(30) Priority: **06.04.1994 US 223810**

(74) Representative:
Watts, Christopher Malcolm Kelway, Dr. et al
AT&T.(UK) Ltd.
5, Mornington Road
Woodford Green Essex, IG8 0TU (GB)

(72) Inventors:
• **Haimi-Cohen, Raziel**
New Jersey 07081 (US)

(54) **Speech recognition system with display for user's confirmation**

(57) A speech recognizer system for use with a telecommunication network wherein an input signal generated onto the network from a first terminal is directed to a speech recognizer for estimating the verbal content of the input signal. The speech recognizer or associated equipment then directs an estimate of the verbal content as an output signal back to the first terminal, the estimate including one or more approximations of the verbal content of the input signal. At the first terminal the user then confirms a correct estimate, or selects from a plurality of approximations, the verbal content of the input signal.



EP 0 676 882 A3



European Patent
Office

EUROPEAN SEARCH REPORT

Application Number
EP 95 30 2109

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int.Cl.6)
A	PATENT ABSTRACTS OF JAPAN vol. 015 no. 209 (E-1072), 28 May 1991 & JP-A-03 058557 (OKI ELECTRIC IND CO LTD) * abstract *	1,5,8,12	
A	DE-A-35 19 915 (TELEFONBAU & NORMALZEIT GMBH) * page 5, line 19 - page 8, line 19; figure 1 *	1,12	
			TECHNICAL FIELDS SEARCHED (Int.Cl.6)
The present search report has been drawn up for all claims			
Place of search THE HAGUE		Date of completion of the search 14 December 1995	Examiner Delangue, P
<p>CATEGORY OF CITED DOCUMENTS</p> <p>X : particularly relevant if taken alone Y : particularly relevant if combined with another document of the same category A : technological background O : non-written disclosure P : intermediate document</p> <p>T : theory or principle underlying the invention E : earlier patent document, but published on, or after the filing date D : document cited in the application L : document cited for other reasons & : member of the same patent family, corresponding document</p>			

EPO FORM 1503 (03.91) (P4/C01)